KATHOLIEKE UNIVERSITEIT LEUVEN

Faculty of Economics and Applied Economics

Department of Economics

Inference for *S*-Gini Poverty Indices

by

Bram THUYSBAERT

Econometrics

Center for Economic Studies
Discussions Paper Series (DPS) 05.09
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

March 2005

DISCUSSION PAPER

Inference for S-Gini poverty indices

Bram Thuysbaert*
K.U. Leuven
March 2005

Abstract

Kakwani [*Econometrica*, 48, 2 (1980), 437-446)] introduced the S-Gini poverty indices as a generalization of Sen's poverty index. I propose a sample estimator for the indices and establish its asymptotic normality under weak conditions. An explicit variance formula is presented. The poverty line is allowed to depend on the income distribution function.

JEL classification: C40; I32
Keywords: poverty; generalized Gini; statistical inference

1 Introduction

In a seminal paper on poverty measurement, Sen (1976) argued that a poverty measure should be sensitive to the inequality among the poor and proposed a poverty index that includes as a component the Gini index of inequality among the poor. A generalization of Sen's poverty index was put forward by Kakwani (1980). Let $F(x)$ be an income distribution function and let $Q(p)$ denote the associated quantile function, i.e. the left-continuous inverse of $F(x)$. Kakwani's poverty indices, which later on were termed S-Gini poverty indices (e.g. Barrett and Donald, 2000), are given by

$$P(s,z) = F(z) - \frac{1}{z}s(s-1)\int_0^1 (1-p)^{s-2} G(pF(z))dp,$$

*I am most grateful to Geert Dhaene for many valuable comments and suggestions. Address for correspondence: Bram Thuysbaert, K.U. Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium. Tel. +32 16 32 68 10. Email: bram.thuysbaert@econ.kuleuven.be

where z is the poverty line and $G(p) = \int_0^p Q(q)dq$ is the generalized Lorenz curve. For $s = 2$, the resulting poverty index is the original measure proposed by Sen. In the remainder of this paper, s is taken to be strictly larger than 1.

Deriving asymptotic distributions and variance formulas for the S-Gini poverty indices has proved to be challenging. The theory of U-statistics is used by Bishop, Formby and Zheng (1997) for the special case of Sen's poverty index. Barrett and Donald (2000) use empirical process theory for the S-Gini inequality and poverty indices. They impose rather restrictive conditions on the income distribution function to ensure weak convergence of the generalized Lorenz process.

Zitikis (2002) has shown that the L-statistic theory of Shorack and Wellner (1986) provides a very convenient tool to analyze the asymptotic behavior of the estimators of the S-Gini inequality indices. In this paper, I use an L-statistic approach to develop a large sample estimation theory for the S-Gini poverty indices. The results are valid under weak conditions and the asymptotic approximations yield variance estimators which have a simple structure and are easy to compute. The poverty line is allowed to depend on the income distribution and thus to be stochastic. None of the above mentioned studies allows for this possibility.

In the next section, I propose an estimator for $P(s, z)$, establish its asymptotic normality and derive an expression for the variance of its limiting distribution. Before concluding, I briefly discuss hypothesis testing.

2 Asymptotic normality

Let X be a random variable representing income and let $F(x)$ be its distribution function. Let $z = z(F)$ be the poverty line. The following assumption is made throughout.

Assumption 1 *F has support on a subset of* $(0, \infty]$.[1]

Using integration by parts and a change of variable, $P(s, z)$ can be written as

$$\begin{aligned} P(s,z) &= F(z) - \frac{s}{zF(z)^{(s-1)}} \int_0^z x[F(z) - F(x)]^{s-1} dF(x) \\ &= F(z) - s\frac{\gamma}{zF(z)^{(s-1)}}, \end{aligned}$$

say. Let $x_1, .., x_n$ be an IID sample of size n from F. Denote the empirical distribution function as $\widehat{F}$ and let $\widehat{z} = z(\widehat{F})$ be the poverty line estimate. The natural

[1]Actually, Assumption 1 could be replaced by the weaker assumption that F has support on a subset of $(a, \infty]$ with finite, possibly negative a. With the lower limit of integration then equal to a, Theorem 1 would continue to hold, though the intermediate result of the Appendix would have to be slightly restated.

estimator of $P(s,z)$ is

$$\widehat{P}(s,\widehat{z}) = \widehat{F}(\widehat{z}) - \frac{s}{\widehat{z}\widehat{F}(\widehat{z})^{(s-1)}} \int_0^{\widehat{z}} x[\widehat{F}(\widehat{z}) - \widehat{F}(x)]^{s-1} d\widehat{F}(x)$$
$$= \widehat{F}(\widehat{z}) - s\frac{\widehat{\gamma}}{\widehat{z}\widehat{F}(\widehat{z})^{(s-1)}}.$$

Observe that, since

$$\widehat{\gamma} = -\frac{1}{s}\int_0^{\widehat{z}} xd[\widehat{F}(\widehat{z}) - \widehat{F}(x)]^s,$$

$\widehat{\gamma}$ can be computed as

$$\widehat{\gamma} = -\frac{1}{s}\sum_{i=1}^{n_{\widehat{z}}} \left[\left(\widehat{F}(\widehat{z}) - \frac{i}{n}\right)^s - \left(\widehat{F}(\widehat{z}) - \frac{i-1}{n}\right)^s\right] x_{(i)},$$

where $x_{(i)}$ is the i-th observation of the ordered sample, $x_{(1)} \leqslant x_{(2)} \leqslant ... \leqslant x_{(n)}$, and $n_{\widehat{z}} = \#\{x_i : x_i \leqslant \widehat{z}; i = 1, ..., n\}$, i.e. the number of observations that are less than or equal to $\widehat{z}$.

Due to the estimation of the poverty line, a few additional assumptions are needed.

Assumption 2 *The estimator $\widehat{z}$ admits the representation*

$$\widehat{z} - z = \int_0^{\infty} \zeta(x) d\widehat{F}(x) + o_p(n^{-1/2})$$

for some function $\zeta(x) : (0,\infty) \to \mathbb{R}$.

Mean and quantile based poverty line estimates can be shown to have this asymptotic representation. For a poverty line equal to a fraction k of mean income μ, $\zeta(x) = k(x - \mu)$. If the poverty line is set to a fraction k of the p-th quantile $Q(p)$, the Bahadur representation (e.g. Ghosh, 1971) yields $\zeta(x) = k\frac{p - I(x \leqslant Q(p))}{f(Q(p))}$, where $I(\cdot)$ denotes the indicator function and $f(Q(p)) = F'(Q(p))$, assumed to be strictly positive here.

Assumption 3 *F is differentiable in a neighborhood of z.*

Assumption 4 *$\int_0^\infty \zeta(x)^2 dF(x) < \infty$.*

Assumptions 2 and 4 imply, trivially, the asymptotic normality of $\sqrt{n}\,(\widehat{z} - z)$.

Lemma 1 *Under Assumptions 1-4,*

$$\widehat{F}(\widehat{z}) - F(z) = \int_0^\infty [I(x \leqslant z) - F(z) + f(z)\zeta(x)]\, d\widehat{F}(x) + o_p(n^{-1/2}),$$

where $f(z) = F'(z)$. *Also,*

$$\sqrt{n}\left(\widehat{F}(\widehat{z}) - F(z)\right) \to_d N(0, \sigma_z^2)$$

with $\sigma_z^2 = \text{var}(I(X \leqslant z) + f(z)\zeta(X))$.

Proof. The result follows from a one-term Taylor expansion. See Zheng (2001) for details. ■

The statistic $\widehat{\gamma}$ is not an L-statistic *stricto sensu*, but it has a very similar structure. Using an L-statistic type of argument, I show in the Appendix that

$$\begin{aligned}
\widehat{\gamma} - \gamma = & \int_0^z \left[\left(\widehat{F}(\widehat{z}) - F(z)\right) - \left(\widehat{F}(x) - F(x)\right)\right](F(z) - F(x))^{s-1} dx \\
& + o_p(n^{-1/2}) \qquad (1) \\
= & \frac{1}{n}\sum_{i=1}^n \int_0^z \{[I(x_i \leqslant z) - F(z) + f(z)\zeta(x_i)] - [I(x_i \leqslant x) - F(x)]\} \\
& \times (F(z) - F(x))^{s-1} dx + o_p(n^{-1/2}).
\end{aligned}$$

The second equality follows from Lemma 1. The difference $\widehat{\gamma} - \gamma$ is now represented as a sample average of IID mean zero variables with finite variance. The Lindeberg-Lévy central limit theorem immediately implies the asymptotic normality of $\sqrt{n}\,(\widehat{\gamma} - \gamma)$. To obtain a convenient variance formula, integrate (1) by parts, yielding

$$\begin{aligned}
\widehat{\gamma} - \gamma = & -\left(\widehat{F}(\widehat{z}) - F(z)\right)\int_0^z x d(F(z) - F(x))^{s-1} \\
& + \int_0^z x\widehat{F}(x) d(F(z) - F(x))^{s-1} \\
& - \int_0^z xF(x) d(F(z) - F(x))^{s-1} \\
& + \int_0^z x(F(z) - F(x))^{s-1} d\left[\widehat{F}(x) - F(x)\right] + o_p(n^{-1/2}),
\end{aligned}$$

and write

$$\begin{aligned}
\int_0^z x\widehat{F}(x) d(F(z) - F(x))^{s-1} &= \int_0^z \left[\int_0^\infty xI(y \leqslant x) d\widehat{F}(y)\right] d(F(z) - F(x))^{s-1} \\
&= \int_0^\infty \left[\int_0^z yI(x \leqslant y) d(F(z) - F(y))^{s-1}\right] d\widehat{F}(x)
\end{aligned}$$

after changing the order of integration. So, using Lemma 1 and rearranging terms,

$$\begin{aligned}\widehat{\gamma}-\gamma = & \int_0^z g(x)d\widehat{F}(x) + F(z)\int_0^z xd(F(z)-F(x))^{s-1} - \int_0^z xF(x)d(F(z)-F(x))^{s-1} \\ & - \int_0^z x(F(z)-F(x))^{s-1}dF(x) + o_p(n^{-1/2})\end{aligned}$$

with

$$\begin{aligned}g(x) = & x[F(z)-F(x)]^{s-1}I(x \leqslant z) + \int_0^z yI(x \leqslant y)d(F(z)-F(y))^{s-1} \\ & - [I(x \leqslant z) + f(z)\zeta(x)]\int_0^z yd(F(z)-F(y))^{s-1}.\end{aligned}$$

It follows that

$$\sqrt{n}\,(\widehat{\gamma}-\gamma) \to_d N(0,\sigma_\gamma^2)$$

with

$$\sigma_\gamma^2 = \mathrm{var}(g(X)).$$

The main result can now be stated.

Theorem 1 *Under Assumptions 1-4 and for $s > 1$,*

$$\sqrt{n}\left[\widehat{P}(s,\widehat{z}) - P(s,z)\right] \to_d N(0,\sigma_P^2),$$

with $\sigma_P^2 = J'\Omega J$, where

$$J = \begin{bmatrix} \frac{s\gamma}{z^2F(z)^{s-1}} \\ 1+\frac{s(s-1)\gamma}{zF(z)^s} \\ -\frac{s}{zF(z)^{s-1}} \end{bmatrix} \quad \textit{and} \quad \Omega = \mathrm{var}\begin{bmatrix} \zeta(X) \\ I(X \leqslant z) + f(z)\zeta(X) \\ g(X) \end{bmatrix}.$$

Proof. The results follows from the joint asymptotic normality of $[\widehat{z},\widehat{F}(\widehat{z}),\widehat{\gamma}]'$ and an application of the delta theorem. ■

If the poverty line is not depending on F, then $\widehat{z} = z$, Assumptions 2 and 4 hold with $\zeta(x) = 0$, Assumption 3 is unnecessary and a simplified variance formula is obtained.

For inference purposes, an estimate of σ_P^2 is needed. A consistent distribution-free estimator of σ_P^2 is obtained by replacing all unknowns by the corresponding sample analogues and the unknown density $f(\cdot)$ by a kernel density estimate $\widehat{f}(\cdot)$. The sample counterpart of $g(x)$ can be computed as

$$\begin{aligned}\widehat{g}(x) = & x[\widehat{F}(\widehat{z}) - \widehat{F}(x)]^{s-1}I(x \leqslant \widehat{z}) + \sum_{i=1}^{n_{\widehat{z}}} d(i)x_{(i)}I(x \leqslant x_{(i)}) \\ & - \left[I(x \leqslant \widehat{z}) + \widehat{f}(\widehat{z})\zeta(x)\right]\sum_{i=1}^{n_{\widehat{z}}} d(i)x_{(i)},\end{aligned}$$

where

$$d(i) = \left(\widehat{F}(\widehat{z}) - \frac{i}{n}\right)^{s-1} - \left(\widehat{F}(\widehat{z}) - \frac{i-1}{n}\right)^{s-1}.$$

The proof of the consistency of the plug-in variance estimator is relatively straightforward and not very insightful. It is therefore omitted.

3 Hypothesis testing

Suppose now that one wants to compare poverty between two income distributions F and G, say. Let P_F and P_G denote the poverty indices associated with F and G respectively. A typical one-sided testing problem would be to test $H_0 : P_F = P_G$ against $H_1 : P_F > P_G$. The natural test statistic is

$$t_{F-G} = \frac{\widehat{P}_F - \widehat{P}_G}{\widehat{\sigma}_{F-G}},$$

where $\widehat{P}_F$ and $\widehat{P}_G$ are the estimates of P_F and P_G and $\widehat{\sigma}^2_{F-G}$ is an estimate of

$$\operatorname{var}(\widehat{P}_F - \widehat{P}_G) = \operatorname{var}(\widehat{P}_F) + \operatorname{var}(\widehat{P}_G) - 2\operatorname{cov}(\widehat{P}_F, \widehat{P}_G).$$

Obviously, if F and G are sampled independently, $\operatorname{cov}(\widehat{P}_F, \widehat{P}_G) = 0$. For dependent samples, the covariance term is nonzero and an expression for the covariance is readily obtained from the joint sample average representations of the constituent parts of $\widehat{P}_F$ and $\widehat{P}_G$ and the delta method. In both cases, if H_0 is true, the test statistic t_{F-G} converges in distribution to a standard normal random variable.

4 Conclusion

Under weak conditions, I demonstrated the asymptotic normality of the S-Gini poverty index estimator. The poverty line was allowed to depend on the income distribution. I also proposed a distribution-free variance estimator. As an alternative to the plug-in variance estimator, resampling methods could be used to estimate the variance of the poverty index estimator. See Xu (2000) for an application of the iterated bootstrap to generalized Gini inequality indices.

Appendix

We have to show that $\Delta = o_p(n^{-1/2})$, where

$$\Delta = \widehat{\gamma} - \gamma - \int_0^z \left[\left(\widehat{F}(\widehat{z}) - F(z) \right) - \left(\widehat{F}(x) - F(x) \right) \right] (F(z) - F(x))^{s-1} dx.$$

Let $J(u) = |u|^{s-1}$ and write

$$\gamma = \int_0^z x J(F(z) - F(x)) dF(x) = \int_0^z x d\psi(F(z) - F(x)),$$

where $\psi(t) = -\int_0^t J(u)du$. Integration by parts gives

$$\gamma = -\int_0^z \psi(F(z) - F(x))dx.$$

Similarly,

$$\widehat{\gamma} = -\int_0^{\widehat{z}} \psi(\widehat{F}(\widehat{z}) - \widehat{F}(x))dx.$$

Now,

$$\begin{aligned} \widehat{\gamma} &= -\int_0^z \psi(\widehat{F}(\widehat{z}) - \widehat{F}(x))dx - \int_z^{\widehat{z}} \psi(\widehat{F}(\widehat{z}) - \widehat{F}(x))dx \\ &= -\int_0^z \psi(\widehat{F}(\widehat{z}) - \widehat{F}(x))dx + o_p(n^{-1/2}). \end{aligned}$$

We obtain

$$\Delta = -\int_0^z \left\{ \frac{\psi(\widehat{F}(\widehat{z}) - \widehat{F}(x)) - \psi(F(z) - F(x))}{A_n} + J(F(z) - F(x)) \right\} A_n dx + R_n,$$

where $A_n = \left(\widehat{F}(\widehat{z}) - F(z) \right) - \left(\widehat{F}(x) - F(x) \right)$ and $R_n = o_p(n^{-1/2})$. Now,

$$\begin{aligned} |\Delta| \leqslant & \sup_x \left| \frac{\psi(\widehat{F}(\widehat{z}) - \widehat{F}(x)) - \psi(F(z) - F(x))}{A_n} + J(F(z) - F(x)) \right| \times \\ & z \left\{ \left| \widehat{F}(\widehat{z}) - F(z) \right| + \sup_x \left| \widehat{F}(x) - F(x) \right| \right\} + |R_n|. \end{aligned}$$

Because of the continuity of $J(u)$ and the Glivenko-Cantelli theorem, observe, after an application of the mean-value theorem, that the supremum factor is $o_p(1)$. Also, both $\left| \widehat{F}(\widehat{z}) - F(z) \right|$ and $\sup_x \left| \widehat{F}(x) - F(x) \right|$ are $O_p(n^{-1/2})$. So, $\Delta = o_p(n^{-1/2})$.

References

Barrett, G. and S. Donald (2000), Statistical inference with generalized Gini indices of inequality and poverty, unpublished manuscript.

Bishop, J. and J. Formby and B. Zheng (1997), Statistical inference and the Sen index of poverty, *International Economic Review*, 38, 381-387.

Ghosh, J.K. (1971), A new proof of the Bahadur representation of quantiles and an application, *Annals of Mathematical Statistics*, 42, 1957-1961.

Kakwani, N. (1980), On a class of poverty measures, *Econometrica*, 48, 427-446.

Sen, A. (1976), Poverty: An ordinal approach to measurement, *Econometrica*, 44, 219-231.

Shorack, G. and J. Wellner (1986), *Empirical processes with applications to statistics*, New York, Wiley.

Xu, K. (2000), Inference for generalized Gini indices using the iterated-bootstrap method, *Journal of Business and Economic Statistics*, 18, 223-227.

Zheng, B. (2001), Statistical inference for poverty measures with relative poverty lines, *Journal of Econometrics*, 101, 337-356.

Zitikis, R. (2002), The asymptotic distribution of the S-Gini index, *Australian and New Zealand Journal of Statistics*, 44, 439-446.